The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation

Bartsch, Lea M ; Singmann, Henrik ; Oberauer, Klaus

Abstract: Refreshing and elaboration are cognitive processes assumed to underlie verbal working-memory maintenance and assumed to support long-term memory formation. Whereas refreshing refers to the attentional focussing on representations, elaboration refers to linking representations in working memory into existing semantic networks. We measured the impact of instructed refreshing and elaboration on working and long-term memory separately, and investigated to what extent both processes are distinct in their contributions to working as well as long-term memory. Compared with a no-processing baseline, immediate memory was improved by repeating the items, but not by refreshing them. There was no credible effect of elaboration on working memory, except when items were repeated at the same time. Long-term memory benefited from elaboration, but not from refreshing the words. The results replicate the long-term memory benefit for elaboration, but do not support its beneficial role for working memory. Further, refreshing preserves immediate memory, but does not improve it beyond the level achieved without any processing.

DOI: https://doi.org/10.3758/s13421-018-0805-9

Originally published at:
Bartsch, Lea M; Singmann, Henrik; Oberauer, Klaus (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. Memory Cognition, 46(5):796-808.
DOI: https://doi.org/10.3758/s13421-018-0805-9

The Effect of Elaboration on Working Memory and Long-term Memory across Age

Lea M. Bartsch & Klaus Oberauer

University of Zurich

and University Research Priority Program “Dynamics of Healthy Aging”

Author Note

We thank Andra Arnicane, Gary Hoppeler and Raffael Schmitt for helping with data collection. The data and the analysis scripts can be accessed in the Open Science Framework (https://osf.io/4n9y3). The project was funded by the University Priority Program Dynamics of Healthy Aging.

Correspondence should be addressed to Lea M. Bartsch, Department of Psychology, Cognitive Psychology Unit, University of Zurich, Binzmuehlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: l.bartsch@psychologie.uzh.ch.

Abstract

Free time to attend to and process information in working memory is key in promoting immediate and delayed retention. One candidate process to cause this benefit is elaboration. We conducted two experiments with young and older adults to investigate whether free time is used for elaboration. Participants remembered lists of nouns, interleaved with short or long free-time intervals, or with filler words connecting all the nouns into a meaningful sentence to assist elaboration. For young adults, assisted elaboration through sentences, and the additional instruction to form a mental image, benefited performance in a working-memory test as much as longer free time, but not more. In contrast, for a delayed test of long-term memory, the benefits of sentence elaboration exceeded those of longer free time. Older adults did not benefit from assisted elaborations in the delayed test, providing further evidence that the long-term memory deficit of older adults arises at least in part from a deficit in elaboration. This elaboration deficit is not driven by a deficit in generating richer representations.

Keywords: elaboration, mental imagery, working memory, long-term memory, aging

The Effect of Elaboration on working memory and long-term memory across Age

In theories of human memory, a distinction is often made between working memory (WM) and long-term memory (LTM). WM is understood as a system that holds mental representations temporarily available for processing, with limited capacity. In LTM information is stored more permanently with presumably unlimited capacity (Cowan, 2008). What is thought to be common to both memory systems is the central role of *control processes* (Atkinson & Shiffrin, 1968). Control processes refer to transient phenomena, which are not permanent features of memory, but which are thought to control what is retained in WM and in LTM. Researchers have tried to isolate these processes, and investigate their underlying mechanisms as well as their benefits for both WM and LTM to this day. One of these processes is called *elaboration,* and it encompasses processes which enrich the memory representation of an item by activating many aspects of its meaning and by linking it into the pre-existing network of semantic associations (Craik & Tulving, 1975; Greene, 1987).

Evidence for the beneficial effects of elaboration on long-term memory is two-fold: First, subjects report to engage in elaborative strategies, such as imagery or sentence generation, spontaneously (Dunlosky & Hertzog, 2001; Richardson, 1998). Second, orienting tasks inducing a richer processing of the memory material have led to better LTM (Craik and Tulving 1975)s. Subsequent work has shown that directly instructing people to engage in elaboration improves their episodic long-term memory (e.g., Bartsch, Singmann, & Oberauer, 2018; Davachi, Maril, & Wagner, 2001). In contrast, the role of elaboration for WM is yet to be determined.

Elaboration as a candidate process in promoting WM

Two findings have given rise to the idea that elaboration might help not only LTM but also WM: (1) A subset of participants report using elaboration during WM tasks, and those who do tend to perform better (Bailey, Dunlosky, & Kane, 2011), and (2) performance in WM tasks is improved by additional free time under certain conditions: (a) if the time is provided between the to-be-remembered items (e.g. Oberauer & Lewandowsky, 2016; Ricker & Hardman, 2017; Tan & Ward, 2008), and (b) more consistently if the items are presented visually rather than auditorily (e.g. Penney, 1975). To date, it is unclear what is causing this effect, but elaboration is a strong candidate: With more free time between the to-be-remembered information, people could engage in this process more. A recent study showed that the beneficial effect of free time between individual memory items is also observed when articulatory rehearsal is blocked, specifically in case of concrete and highly imaginable words, which are in general easier to elaborate (Souza & Oberauer, 2018). This result suggests that subjects might have engaged in a form of elaboration, but as of now, there is no direct evidence to support this assumption.

Recent studies investigating the effect of experimentally inducing elaboration have challenged the claim of a beneficial effect of elaboration on WM: Instructing participants to form a vivid mental image of parts of the memoranda after list presentation did not benefit WM, although it improved LTM (Bartsch, Loaiza, Jäncke, Oberauer, & Lewis-Peacock, 2019; Bartsch & Oberauer, 2019; Bartsch et al., 2018). There might be a reason why the elaboration instruction in our earlier studies did not improve WM: Asking participants to elaborate a set of words after an entire word list has been encoded into WM might make it too hard for them to access the words they should elaborate. Elaboration might be easier when it can occur in between presentation of individual items. This assumption receives further plausibility by the fact that free time improves WM only when added in between list items, not when added at the end of the list (Oberauer & Lewandowsky, 2016).

The current study was designed to make elaboration as easy as possible for participants. To this end, we provided the enriching information via sentences rather than asking participants to generate the enriched representations themselves. The technique of forming sentences of individual to-be-remembered words has previously been reported in strategy assessment and training studies (McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). For instance, in a strategy assessment study by Bailey and colleague's, subjects reported to having formed sentences or mental images in 14% and 8% of the trials of a complex span task, respectively, making these the most frequently reported strategies after rote repetition and reading. (Bailey, Dunlosky, & Hertzog, 2009; see also Dunlosky & Kane, 2007). Furthermore, the degree to which an individual reportedly used these two elaborative strategies correlated with their WM span performance.

An early study investigating the effect of sentences generation and mental imagery on long-term memory recognition of paired associates showed that the visualization of a mental image led to better recall than simply being presented with the word pair embedded in a sentence (Bower & Winzenz, 1970). This finding suggests that embedding to-be-remembered words in sentences helps people to elaborate because they can draw on the scene or image described in the sentence rather than having to invent a mental image on the spot. Participants might obtain a larger benefit from elaboration – not only for LTM but also for WM – if, in addition to presenting a sentence, mental imagery of the sentence's meaning is encouraged by the instruction.

Against the possibility that elaboration helps WM stands a long line of research has shown that carrying out a secondary task that involves processing additional material during encoding or maintenance impairs immediate memory performance (Oberauer et al., 2018). Thereby, the strategy of forming sentences and creating visual images – whether on one's own initiative or in response to an experimenter's instruction – could result in interference in WM. Taken together there are, a priori, good reasons for predicting a beneficial as well as a detrimental effect of elaboration on WM.

The first goal of the present study is to investigate whether elaboration of information in WM, when facilitated by providing enriching information, improves not only long-term retention but also performance in an immediate test of WM.

Age-related shortfalls in elaboration processes relate to LTM deficits

Both WM and LTM decline in old age. The LTM deficit of older adults can in part be attributed to a deficient effectiveness of elaboration in older compared to young adults (Bartsch, Loaiza, Jäncke, et al., 2019; Bartsch & Oberauer, 2019). More precisely, we have previously shown that older adults did not benefit from elaboration in LTM although their brain activation patterns during the formation of mental images at encoding were differentiable from a repeated reading condition – similar to young adults (Bartsch, Loaiza, Jäncke, et al., 2019).

Smith (1980) has argued that the reason why older adults' LTM does not benefit from elaboration is that they have difficulties generating the necessary enrichment of the learning material. This *generation-deficit hypothesis* states that older adults exhibit smaller elaboration benefits on LTM, compared to young adults, when they have to generate the richer representations themselves. In line with this claim, Rankin and Collins (1985) provided evidence that older adults' memory benefited when elaborations were given to them in the form of sentences equivalently to young adults, but were less likely than young adults to generate relevant elaborations themselves. Similarly, Cherry (1993) showed that explanatory elaborations provided in the form of sentences at encoding enhanced delayed memory for target adjectives in young and older adults, but only when the elaborations were given again at retrieval.Age-related production deficits for encoding strategies have been reported more generally as well: for instance, older adults are less likely than young adults to spontaneously use effective strategies when studying paired associates, but they can successfully use them if instructed to (Dunlosky & Hertzog, 2001; see Kausler, 1994, for a review).

In contrast to the above LTM related production deficits, no such age-differences in the proportion of normatively effective strategies occur in WM span tasks (Bailey et al., 2009). Both older and young adults reported using mental imagery (OA: 11 %, YA: 14%), and sentences (OA: 12 %, YA: 8%) comparatively frequently across the trials of two WM span tasks. Although the use of these normatively effective strategies accounted for individual differences in WM span performance in that study in general, it did not account for age-related variance in WM span performance (Bailey et al., 2009). Experimental investigations of the effect of elaboration on WM performance showed no age-difference in the null-effect of elaboration: The instruction to form a vivid mental image of parts of the memoranda after list presentation did not benefit WM, in neither young nor older adults (Bartsch, Loaiza, Jäncke, et al., 2019; Bartsch & Oberauer, 2019).

Taken together, LTM deficits in older adults might at least in part arise from a deficiency in generating elaborations at encoding. Therefore, the second goal of the present study was to investigate why older adults show so little benefit of elaboration, and specifically, to test the generation-deficit hypothesis. If this hypothesis is correct, then providing older adults with sentences that enrich the to-be-remembered words should help them overcome the generation deficit, so that their LTM benefits from elaboration as much as young adults do. Furthermore, extending the first goal of the experiment, we aimed to investigate whether elaboration of information in WM, when facilitated by providing enriching information, improves WM performance in older adults as well.

The present study

The first goal of the present study is to investigate whether young and older adults benefit from assisted elaboration through sentences in a test of WM. We therefore tested immediate serial-recall performance for nouns that were interleaved by filler words embedding them into meaningful sentences. This *assisted-elaboration* condition was compared to two baseline conditions without interleaved words, which differed in the amount of free time in between the nouns. In the *short baseline* condition the time for presenting the interleaved words was cut out, whereas in the *long baseline* condition these time intervals were retained as free time. We expected that the longer free time should lead to better serial recall in the long than the short baseline (i.e. a free time benefit). If spontaneous elaboration underlies this free-time benefit for serial-order WM, we predict that immediate serial recall in the sentence condition is as good as in the *long baseline* condition, or even better due to the external assistance of elaboration.

We further included a delayed memory test as a manipulation check: The extent to which our manipulation of elaboration increases the degree to which people engage in elaboration (relative to the other conditions) should be reflected in an increase of LTM, at least in young adults. Furthermore, if spontaneous elaboration underlies the free-time benefit in WM, the effect of our experimental manipulations on delayed recall should mirror those on the WM test.

Specifically, if providing sentences induces elaboration that is about as effective as people's spontaneous elaboration, the *long baseline* and sentence condition should show an equivalent beneficial effect in comparison to the short baseline in the delayed memory test as well; if providing sentences facilitates even more effective elaboration, then performance should exceed that in the long baseline condition for both the WM and the LTM tests. Conversely, if spontaneous elaboration is not responsible for the free-time benefit in WM, then assisting elaboration by providing sentences should improve LTM but not WM relative to the *long baseline*.

Our second goal is to test the hypothesis that the LTM deficit in older adults arises in part from a deficit in generating the enrichment of the memory material necessary for effective elaboration. To this end we tested episodic LTM through a delayed free-recall test of the nouns that had been presented as memoranda for the immediate serial-recall trials. On the generation-deficit hypothesis we expect that, when older adults are provided with sentences that enrich the memory material for them, their LTM should benefit from elaboration to a comparable degree as that of young adults.

Method

Materials and Procedure

In the two experiments presented here, we asked participants to remember short lists of nouns in serial order. The stimuli were drawn from a pool of 450 German concrete nouns. The nouns were between three and nine letters long and had a mean normalized lemma frequency of 23.41/million (drawn from the dlexdb.de lexical database).

The sequence of an experimental trial is illustrated in Figure 1. The presentation rate of each word was self-adjusted by each participant at the beginning of the experiment: Example sentences were presented word-by-word centrally on the screen, and subjects were asked to adjust the presentation time so they were comfortably able to read the sentences. In the main experiment, participants were informed prior to each trial about the following experimental condition. After sequential presentation of the to-be-remembered words, a WM test followed immediately: An array of words was displayed, and participants were to reconstruct the memory list by clicking on the list words in their serial order. The response options in the array consisted of all of the words of that trial's memory list, and the same number of new items. The position of the options on the screen was random, and participants used the mouse to select among them at their own pace.

There were four encoding conditions. In the *short baseline* condition, the to-be-remembered words were presented individually in the center of the screen for the amount of time taken from the adjustment period (e.g. 500 ms), interleaved by a short ISI of 100 ms. In the *long baseline* each word was followed by the presentation of a blank screen for two times the word presentation time (e.g. 1000 ms), equivalent to the time to present two filler words. For the *sentence grammar* and *sentence imagery* conditions, the to-be-remembered words were presented within a sentence, each word followed by, on average, two filler words. The nouns were presented in bold, and fillers were never nouns, thereby making the memoranda very distinct to the participant. Five independent subjects were invited to the lab to create the sentences, and the first author selected 75 sentences to be used in the experiments, based on their meaningfulness and in accordance with the desired 2:1 ratio of filler words to target nouns. In both sentence conditions, all the words were presented centrally on the screen at the pace adjusted by each individual in the beginning of the experiment. In the *sentence grammar* condition, the subjects were asked to judge whether the sentence was grammatically correct, which they were in 50% of the cases. In the *sentence imagery* condition, the subjects were asked to form a vivid mental image of the scene described in the sentence. Following the memory test, the subjects were to rate the vividness of the created mental image (*sentence imagery* condition), the grammatical correctness of the sentence (*sentence grammar* condition), or how well they were able to read the words (*short* & *long baseline* condition).

There were four trials of the WM task per block, one of each condition. The experiment comprised eight blocks. An unrelated distracter task followed each block, in which the participants had to indicate the correctness of visually presented math equations (e.g. 9 x 8 = 72) for 2 minutes. After that followed a typed delayed free recall memory test, wherein the participants were asked to recall as many memory items from the previous block as possible. This test served to assess the effect of each experimental condition on episodic LTM. The participants were made aware of the delayed memory test before the start of the experiment.

We carried out two nearly identical experiments. In Experiment 1, both age groups did the immediate serial recall task with a memory set size of five nouns. In this way we could compare performance of the two age groups at the same level of load on WM. The same nominal memory load, however, poses a higher demand on WM for old adults who have, on average, a lower WM capacity than the young. This could impair their ability to form a robust, accessible trace of the memoranda in episodic LTM, and in particular it could compromise their ability to form an integrated representation of the meaning of the given sentence, thereby undercutting any potential benefit of elaboration we aimed to induce. Therefore, we ran Experiment 2, which was identical to Experiment 1, but varied memory set size between the age-groups (6 nouns for young vs. 4 nouns for older adults). The reduced memory load for old adults served to compensate for their reduced WM capacity.

Data Analysis

We analyzed Experiments 1 and 2 using a Bayesian generalized linear mixed model (BGLMM) implemented in the R package *brms* (Bürkner, 2017, 2018). For the data of the WM task the dependent variable was the correctness (0 or 1) of each of the responses in each trial (serial recall of first to last item) of each condition per participant. Correct responses were defined as choosing the target item from the alternatives (i.e., all other list items and new items). Therefore, we assumed a Bernoulli data distribution predicted by a linear model through a logit link function (i.e., a repeated-measures logistic regression).

For the data of the LTM task the dependent variable was the proportion of correctly recalled items in each condition per block and participant. Correct responses were defined as recalling one of the nouns that was presented in the previous block of WM trials. Here, we assumed a Gaussian data distribution predicted by a linear model through an identity link function.

For both analyses, the fixed effects were condition and age group as well as their interaction. Following the recommendation of Barr and colleagues (Barr, Levy, Scheepers, & Tily, 2013; see also Schielzeth & Forstmeier, 2009) we implemented the maximal random-effects structure justified by the design, including a by-participant random intercept and a random slope for condition. In addition, we estimated the correlation among the random-effects parameters.

The regression coefficients were given weakly informative Normal priors with a mean of 0 and the standard deviation of 5. We used completely non-informative priors for the correlation matrices, so-called LKJ priors with shape parameter 1. Bayesian procedures provide posterior probability distributions of the model parameters (i.e., the regression weights) that express uncertainty about the estimated parameters. The highest density regions (HDRs) of these posteriors can be used for statistical inference. A 95% HDR represents the range in which the true value of a parameter lies with probability 0.95, given model and data (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). If zero lies outside the Bayesian HDR there is strong evidence for the existence of the corresponding effect; although the strength of evidence varies continuously, for simplicity we will describe effects as "credible" if their HDRs exclude zero.

We used an MCMC algorithm (implemented in Stan; Carpenter et al., 2017) that estimated the posteriors by sampling parameter values proportional to the product of prior and likelihood. These samples are generated through 4 independent Markov chains, with 500 warmup samples each, followed by 1000 samples drawn from the posterior distribution which were retained for analysis. Following Gelman and colleagues (2013), we confirmed that the 4 chains converged to the same posterior distribution by verifying that the $\hat{R}$ statistic – reflecting the ratio of between-chain variance to within-chain variance – was < 1.02 for all parameters, and we visually inspected the chains for convergence.

Because we use Bayesian statistics, we did not use power considerations for deciding on our sample size – the concept of power is not defined in Bayesian statistics. What best corresponds to it is the precision of the posterior estimates of standardized effect sizes. We based our sample sizes on our previous studies on this topic (Bartsch, Loaiza, Jäncke, et al., 2019;

Bartsch & Oberauer, 2019), which yielded sufficiently precise posteriors to permit confident inferences.

Participants

For Experiment 1 we recruited 24 students (13 female) from the University of Zurich and 24 healthy older adults (14 female) from the Zurich community as participants. Participants in Experiment 2 were 21 students (13 female) from the University of Zurich and 20 healthy older adults (10 female) from the Zurich community. In both studies, participants were compensated with either 15 Swiss Francs (about 15 USD) or partial course credit for the one-hour experiment. The studies were carried out in agreement with the rules of the Ethics Committee of the Faculty of Arts and Sciences of the University of Zurich.

Cognitive functioning was screened with the MMSE (Mini-Mental Status Examination; Folstein, Folstein, & McHugh, 1975), indicating age-typical cognitive abilities in our sample of older adults (Experiment 1: M = 29, SD = 1.14, range = 26 – 30; Experiment 2: M = 28.86, SD = 1.38, range = 25 - 30). Table 1 shows the descriptive statistics and posterior distributions of the age effects of our sample. The evidence indicates slower processing speed in the older compared with the young adults as measured by the digit-symbol test (Petermann & Wechsler, 2012). The older adults showed better performance than the young adults in a computerized vocabulary test (Mehrfachwahl-Wortschatz Test Version B, Lehrl, 2005), consisting of 37 items in which participants are supposed to find an existing word among four similarly sounding non-words. The MWT-B is a marker test for crystallized intelligence. Hence, our sample of young and old adults showed typical age differences in processing speed and measures of crystallized intelligence (Li et al., 2004).

Results

To draw inferences about the effect of elaboration and free time on WM as well as these processes' impact on LTM formation, we first focus on the results of Experiments 1 and 2 separately. We subsequently evaluate these results in a joint analysis over both experiments, including memory load as an additional factor. All data and analysis scripts can be assessed on the Open Science Framework (https://osf.io/4n9y3).

Experiment 1

The mean self-chosen reading speed was 481ms (SD = 145) per word for the young adults and 604ms (SD = 113) for the older adults.

Manipulation Check

Figure 3 shows the mean free recall performance and their corresponding 95% within-subject confidence intervals in the long-term memory task of Experiment 1. The posterior effect estimates are presented in Table 3. The first comparison of interest concerned our manipulation check, assessing to what extent our manipulation of elaboration increased the degree to which people engage in elaboration. We therefore compared performance in the sentence imagery condition to both baselines. As seen in Table 3, for young adults the sentence imagery condition yielded better performance than both baseline conditions, demonstrating that assisted elaboration effectively boosted episodic LTM, even more so than the long baseline. Moreover, the younger adults showed higher delayed recall performance in the sentence imagery condition than the sentence grammar condition, suggesting that forming an image of the sentence's meaning was a more effective elaboration process than evaluating the sentence's grammaticality. In conclusion, this manipulation check was successful, and we ensured, that our assisted-elaboration manipulation increased the amount or effectiveness of elaboration beyond that achieved in the long baseline.

Does elaboration improve WM?

Figure 2 shows the mean serial-recall performance and their corresponding 95% within-subject confidence intervals in the working-memory task of Experiment 1. The posterior effect estimates are presented in Table 2. There was a main effect of age, with younger adults outperforming the older adults in the WM task across conditions (older vs. young Δ = -0.88, 95% HDR = [-1.33, -0.44]). Our first question was whether our manipulation of free time between the presentation of items in a memory list replicated the usual effect on immediate serial recall (WM task). There was a credible difference between the short and long baseline for both young and older adults, implying that participants had better memory for items interleaved with a longer free-time interval than for items without this free time (see Figure 2 and Table 2). Next, we were interested how the sentence conditions affected performance compared to the two baselines. The BGLMM revealed that performance of young adults in the sentence-imagery condition approximated that in the long baseline. Yet, the comparison to the short baseline revealed that performance in the sentence imagery condition was not credibly better either.

The older adults, by contrast, performed more poorly in the sentence-imagery condition than the long baseline. Immediate serial recall was poorer in the sentence-grammar condition than the long baseline for both age groups, suggesting that the process of evaluating the sentence's grammaticality was not an effective form of elaboration for WM.

Does providing enriched information overcome older adults' elaboration deficit?

For the second goal of our study, we then looked at the effects of age on the delayed memory data. As seen in Figure 3 and supported by evidence for a main effect of age (older vs. young: $\Delta = -0.15$, 95% HDR = [-0.23, -0.06]) the older adults remembered less words than the younger adults. Their performance was equivalent across all conditions, including assisted elaboration through sentences. Figure 6 depicts the interaction effect of condition by age group, showing evidence that the beneficial effect of sentence imagery – in comparison to the two baselines and to the sentence-grammar condition – was larger for young than for old adults.

Experiment 2

The mean self-chosen reading speed was 462.5ms (SD = 135.23) per word for the young adults and 656.84ms (SD = 115.29) for the older adults.

Manipulation Check

Figure 5 shows the mean free recall performance and their corresponding 95% within-subject confidence intervals in the long-term memory task of Experiment 2. The posterior effect estimates of the BGLMM are presented in Table 5. For young adults, episodic LTM in the sentence-imaging conditions surpassed that in the long baseline. Moreover, the younger adults showed higher delayed recall performance in the sentence imagery condition than the sentence grammar condition, confirming that forming a mental image is the more effective form of elaboration.

Does elaboration improve WM?

Figure 4 shows the mean serial recall performance and their corresponding 95% within-subject confidence intervals in the working memory task of Experiment 2. The posterior effect estimates are presented in Table 4. As the older adults were presented with only four words, they now outperformed the younger adults in the WM task across conditions (older vs. young Δ = 1.13, 95% HDR = [0.64, 1.64]).

Again, our first question was how the manipulation of free time between the presentation of items in a memory list affected immediate serial-recall performance. The analysis yielded a credible benefit of longer free time only for the young adults; old adults' performance might have been too close to ceiling to allow a sizeable free-time benefit (see Figure 4 and Table 4).

Next, we were interested how the sentence conditions affected performance compared to the two baselines. The BGLMM revealed that for young adults the performance in the sentence imagery condition again approximated that of the long baseline – in contrast to the LTM data, where assisted elaboration resulted in performance surpassing that of the long baseline.

Yet, as in Experiment 1, the comparison to the short baseline revealed that performance in the sentence imagery condition was not credibly better either. Again, the older adults did not show this pattern and instead performed more poorly in the sentence grammar and sentence imagery conditions than in the long baseline, and even somewhat worse than the short baseline. As in Experiment 1, immediate serial recall was poorer in the sentence-grammar condition than the long baseline for both age groups.

Does providing enriched information overcome older adults' elaboration deficit?

Now that overall performance in the WM task was higher for the older adults than the young adults, thereby more than compensating for the age deficits at the WM stage, old adults should have been able to fully process the sentences provided, thereby making optimal use of the assisted elaboration. Figure 5 shows that, whereas young adults again benefited from elaboration, older adults did not.

As seen Figure 5 and supported by evidence for a main effect of age (older vs. young: $\Delta = -0.10$, 95% HDR = [-0.17, -0.03]) the older adults remembered a lower proportion of words than the younger adults, and their performance was equivalent across all conditions, including assisted elaboration through sentences. Figure 7 depicts the interaction effect of condition by age group, showing evidence that, relative to both baselines, sentence imagery improved episodic memory more for young than for old adults.

Joint Analysis

To investigate how memory load affects the effects of elaboration and free time across both experiments, we jointly analyzed the data of both experiments with memory load as an addition factor, in another BGLMM.

Working Memory

The analysis confirmed the expected effect of memory load ($\Delta = -1.00$, 95% HDR = [-1.34, -0.68]), with lower performance at higher load. Next, we were interested in whether the effects of our manipulation of encoding conditions varied with memory load. The analysis of the posterior effect estimates of the effect of memory load across conditions revealed that there was no evidence for any such interaction (see Table 6). We therefore collapsed the analysis over memory load. The condition contrast posteriors collapsed over memory load are shown in Table 7. With the power of the joint data from Experiment 1 and 2, these results reveal credible evidence for a difference between performance in the sentence imagery and short baseline in young adults.

Long-Term memory

The analysis revealed that performance in the delayed memory task was unaffected by memory load (high vs. low: Δ = -0.03, 95% HDR = [-0.09, 0.03]). As seen in Table 8, mirroring the analysis for WM, memory load did not interact with any of the conditions in LTM. We therefore collapsed the analysis over memory load. The condition contrast posteriors collapsed over memory load are shown in Table 9. The contrast of short compared to long baseline revealed evidence for a credible difference, indicating a beneficial effect of free time also for delayed memory in young adults. Yet again, the sentence conditions credibly differed from the long baseline in young adults.

Discussion

The first goal of the present study was to examine to what extent young and older adults can benefit from assisted elaboration through sentences in a test of WM as well as a test of LTM, and to study elaboration as a potential cause for the free-time benefit in WM. We assumed, that if free time is naturally used to engage in elaboration, then assisted elaboration should add little (if anything) to both WM and LTM in young adults - who we assume are good at elaborating - and more for old adults who have been shown to be deficient in generating elaborations themselves.

Elaboration and the free time benefit on working memory

Our results confirm that young adults benefit from free time interleaving the to-be-remembered items in WM, and so do older adults when performance is not close to ceiling. One potential explanation for this effect is that people use free time to elaborate the memoranda, and that helps immediate recall.

We know from self-report studies that only one fourth of the subjects indicate to spontaneously elaborate in WM tasks (e.g. Bailey et al., 2011; Dunlosky & Kane, 2007). We therefore expected that instructing *all* participants to form a mental image of the memoranda should have led to more consistent elaboration than the long baseline. Additionally, the fact that we also assisted elaboration by providing meaningful sentences should have boosted the effectiveness of elaboration. The effects of this manipulation on LTM in the young-adult group corroborated that assumption: Both sentence conditions resulted in a substantial benefit for LTM compared to the short baseline, and in case of the sentence imagery condition also compared to the long baseline. This implies that, compared to the long baseline condition, in the sentence imagery condition more people engaged in elaboration, or they did so more effectively. These findings demonstrate that we managed to increase the amount or effectiveness of elaboration, at least in young adults. Then the critical question was, how this affected participants' WM. We found that the sentence imagery condition never surpassed the long baseline in young adults (and in fact always ended up a bit short). The older adults, instead of being able to compensate their elaboration deficit, performed more poorly with assisted elaboration than in the long baseline. These two findings question the idea that the free-time benefit on WM is to a large extent due to elaboration.

Therefore, the WM results are better explained by the following interpretation: The sentence imagery condition enabled young adults to create durable representations, with deeper associations within LTM, which led to better memory for those words in the LTM test. In line with previous research showing that subjects can flexibly use their LTM in WM tasks (e.g.

Lewis-Peacock & Postle, 2008; Schurgin, Cunningham, Egeth, Brady, & Hall, 2018), they could also draw on these deeper associations during the WM test, thereby improving performance to a level approximating that of the long-baseline condition. However, in the long baseline another process strengthened WM, a process that did not promote more durable or more accessible LTM representations. In older adults, elaboration is less effective, and therefore their WM performance did not reach that of the long-baseline condition.

An alternative explanation is that elaboration did improve both WM and LTM in our study, but that the beneficial effect in WM is counteracted by a secondary-task load. More precisely, the enriched representations created by reading sentences and forming mental images could have created interference in WM in the same way as it is the case for sentence reading in reading span tasks. Elaboration – whether it is experimentally induced or initiated spontaneously -- imposes a secondary task demand, and immediate serial recall is known to be vulnerable to secondary tasks (e.g. Jonker & Macleod, 2015, see Oberauer et al., 2018 for an overview (benchmark 5.2.)). Older adults were more strongly affected by the secondary task demand of elaboration, thereby showing poorer WM performance in the sentence imagery compared to the long baseline (Clapp & Gazzaley, 2012; Hasher, Zacks, & May, 1999). This alternative explanation assumes that elaboration has a beneficial effect on WM representations, but the conclusion with regard to the question we started from remains unchanged: Elaboration has no beneficial net effect for WM, and therefore the free-time benefit cannot be explained as an effect of elaboration.

The role of mental imagery for elaboration

Our results show that the semantic context provided by the sentences alone had only a modest beneficial effect on LTM, and no credible effect at all on WM. Only the additional instruction to form a mental image resulted in WM performance approximating that of the long baseline, and also promoted the largest LTM effect. Following the dual coding theory, people store associations between two types of information, verbal and visual, separately in LTM (Paivio, 1991). In this way, adding mental imagery to sentence reading results in more retrieval cues than sentence reading alone, which could promote also better immediate recall in a WM task. In line with this claim, a recent review put forward – based on evidence from the imagery literature – that people can use at least two forms of mental representations, verbal as well as depictive representations (Pearson & Kosslyn, 2015). These authors further argue that images contain much implicit information, which makes such depictive representations especially useful for memory. Indeed, evidence suggests an overlap of mental imagery and visual working memory: both share neural correlates and mechanisms in the sensory cortex (e.g. Albers, Kok, Toni, Dijkerman, & De Lange, 2013; see Pearson, Naselaris, Holmes, & Kosslyn, 2015 for an overview), and individuals with higher sensory strength of mental images have been shown to rely more on imagery as a mnemonic strategy in visual WM tasks (Borst, Ganis, Thompson, & Kosslyn, 2012). The present results extend the evidence for the possible beneficial role of mental imagery to a verbal memory task. This conclusion stands in contrast to previous result showing that the correlation between mental imagery strength and individual's WM performance does not hold for verbal tasks (Keogh & Pearson, 2014). This discrepancy can be explained by the nature of the imagery condition in the present study: The sentences provided already a more imaginable representation of the verbal material and thereby aided the formation of a mental image, so that the impact of individual differences in the ability to create images was diminished.

Taken together, assisted elaboration through sentences and mental imagery instruction resulted in a larger memory benefit in both immediate and delayed tests compared to merely integrating the to-be-remembered words into a meaningful sentence, attesting to the important role of mental imagery for elaboration.

Older adults' LTM deficit relates to elaboration

As indicated by previous research, older adults present a specific deficit in the effectiveness of elaboration that has been argued to contribute to their decline in LTM. Here, older adults were assisted in two ways: First, elaboration was assisted by the sentences provided, thereby controlling for any deficit in *generating* richer representations and/or associations on the spot. Second, by reducing the memory load at encoding to account for any initial deficits in Experiment 2, we made it easier for older adults to fully process and mentally integrate the sentences. Nevertheless, across both experiments, older adults' episodic LTM did not benefit from either sentence condition compared to the baselines, in contrast to the young adults. The lack of a benefit of the sentence imagery compared to the sentence grammar condition is in line with previous work revealing a reduced benefit of mental imagery strategy use with aging (e.g., Kemp & Newson, 2005; Palladino & De Beni, 2003). Our finding of no benefit of assisted elaboration in old adults contradicts the generation-deficit hypothesis (Smith, 1980), and also rules out the assumption that WM capacity constraints are responsible for older adults' ineffective elaboration.

Our findings disagree with some previous evidence in the field (e.g. Rankin & Collins, 1985) showing that older adults can benefit from elaboration, if richer representations are provided. In contrast to having embedded all to-be-remembered words of a list within one sentence in the present study, Rankin and Collins provided individual sentences per target word. It could be that this provides a more effective form of elaboration for old adults, though we cannot think of a reason why that should be the case.

Another explanation for the lack of an elaboration benefit in older adults could be, that they would have needed more time to process the sentences and to form a mental image. Past research on another form of LTM (associative memory) has shown that time to process information at encoding improves LTM of old adults (Bartsch, Loaiza, & Oberauer, 2019). That said, as subjects adjusted the presentation time at the beginning of the experiment to their personal reading speed, we adjusted for individual differences in speed of sentence processing. Furthermore, in a recent study by Hinault and colleagues, older adults were given 8 seconds to encode word pairs and form an interactive mental image (compared to 6 seconds in young adults), yet the beneficial effect of elaboration – compared to rote rehearsal – was much larger in young than older adults (Hinault, Lemaire, & Touron, 2017).

The present results also add to our previous findings showing that mental imagery instructions resulted in differentiable brain activation patterns in older adults compared to repeated reading and refreshing – demonstrating that old adults followed the elaboration instruction --, but in contrast to younger adults, this did not result in a LTM benefit (Bartsch, Loaiza, Jäncke, et al., 2019). The present study thereby provides further evidence that older adults' LTM deficit arises at least in part from a deficit in elaboration, and adds to the literature that this deficit does not arise from an inability to generate the enriched representations needed for elaboration. Rather, enriched representations fail to improve accessibility of episodic memory traces in older adults.

Conclusion

The present study showed that elaboration through embedding words in sentences and mental imagery improves LTM and, to some extent, also WM. However, elaboration is not underlying the free time benefit for tests of WM. Furthermore, we add to the evidence that enriched representations fail to improve accessibility of episodic memory traces in older adults, leading to the pronounced age-related LTM deficit.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier. https://doi.org/10.1016/S0079-7421(08)60422-3

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging*, *24*(1), 82–92. https://doi.org/10.1037/a0014078

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition*, *39*(3), 447–461. https://doi.org/10.3758/s13421-010-0034-3

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartsch, L. M., Loaiza, V. M., Jäncke, L., Oberauer, K., & Lewis-Peacock, J. A. (2019). Dissociating refreshing and elaboration and their impacts on memory. *NeuroImage*, *199*, 585–597. https://doi.org/10.1016/j.neuroimage.2019.06.028

Bartsch, L. M., Loaiza, V. M. V. M., & Oberauer, K. (2019). Does Limited Working Memory Capacity Underlie Age Differences in Associative Long-Term Memory? *Psychology and Aging*, *34*(2), 268–281. https://doi.org/http://dx.doi.org/10.1037/pag0000317

Bartsch, L. M., & Oberauer, K. (2019). Adult age differences in refreshing and elaboration and their consequences for working memory and long-term memory. *PsyArXiv*. https://doi.org/10.31234/OSF.IO/Q76FY

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance , and their contributions to long-term memory formation. *Memory & Cognition*, *46*(5). https://doi.org/10.3758/s13421-018-0805-9

Borst, G., Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2012). Representations in mental imagery and working memory: Evidence from different types of visual masks. *Memory & Cognition*, *40*(2), 204–217. https://doi.org/10.3758/s13421-011-0143-7

Bower, G. H., & Winzenz, D. (1970). Comparison of associative learning strategies. *Psychonomic Science*, *20*(2), 119–120. https://doi.org/10.3758/BF03335632

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*, *10*(1), 395–411. https://doi.org/10.32614/rj-2018-017

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). *Stan* : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Clapp, W. C., & Gazzaley, A. (2012). Distinct mechanisms for the impact of distraction and interruption on working memory in aging. *Neurobiology of Aging*, *33*(1), 134–148. https://doi.org/10.1016/J.NEUROBIOLAGING.2010.01.012

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*(07), 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9.What

Craik, F. I. M., & Tulving, E. (1975). Depth of Processing and the Retention of Words in Eisodic Memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294.

Davachi, L., Maril, a, & Wagner, a D. (2001). When keeping in mind supports later bringing to mind: neural markers of phonological rehearsal predict subsequent remembering. *Journal of Cognitive Neuroscience*, *13*(8), 1059–1070. https://doi.org/10.1162/089892901753294356

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, *29*(2), 247–253. https://doi.org/10.3758/BF03194918

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, *60*(9), 1227–1245. https://doi.org/10.1080/17470210600926075

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis, 3rd edition. Chapman & Hall/CRC.

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin*, *102*(3), 403–413. https://doi.org/10.1037/0033-2909.102.3.403

Hasher, L., Zacks, R. T., & May, C. P. (1999). *Inhibitory control, circadian arousal, and age*. *Attention and Performance* (Vol. 17).

Hinault, T., Lemaire, P., & Touron, D. (2017). Aging effects in sequential modulations of poorer-strategy effects during execution of memory strategies. *Memory*, *25*(2), 176–186. https://doi.org/10.1080/09658211.2016.1146300

Jonker, T. R., & Macleod, C. M. (2015). Disruption of relational processing underlies poor memory for order. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *41*(3), 831–840. https://doi.org/http://dx.doi.org/10.1037/xlm0000069

Kausler, D. H. (1994). *Learning and memory in normal aging.* Academic Press.

Keogh, R., & Pearson, J. (2014). The sensory strength of voluntary visual imagery predicts visual working memory capacity. *Journal of Vision*, *14*(12), 7.

Lewis-Peacock, J. A., Drysdale, A. T., & Postle, B. R. (2015). Neural evidence for the flexible control of mental representations. *Cerebral Cortex*, *25*(10), 3303–3313. https://doi.org/10.1093/cercor/bhu130

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary Activation of Long-Term Memory Supports Working Memory, *28*(35), 8765–8771. https://doi.org/10.1523/JNEUROSCI.1953-08.2008.Temporary

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory and Cognition*, *29*(1), 10–17. https://doi.org/10.3758/BF03195736

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Oberauer, K., & Lewandowsky, S. (2016). Control of information in working memory: Encoding and removal of distractors in the complex-span paradigm. *Cognition*, *156*, 106–128.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., … Hurlstone, M. J. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *45*(3), 255.

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proceedings of the National Academy of Sciences of the United States of*

America, *112*(33), 10089–10092. https://doi.org/10.1073/pnas.1504933112

Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: functional mechanisms and clinical applications. *Trends in Cognitive Sciences*, *19*(10), 590–602.

Penney, C. G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin*, *82*(1), 68.

Petermann, F., & Wechsler, D. (2012). *Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV)*. Frankfurt am Main: Pearson.

Rankin, J. L., & Collins, M. (1985). Adult age differences in memory elaboration. *Journal of Gerontology*, *40*(4), 451–458. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4008880

Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*(4), 597–614.

Ricker, T. J., & Hardman, K. O. (2017). The Nature of Short-Term Consolidation in Visual Working Memory. https://doi.org/10.1037/xge0000346

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Schurgin, M. W., Cunningham, C. A., Egeth, H. E., Brady, T. F., & Hall, M. (2018). Visual Long-term Memory Can Replace Active Maintenance in Visual Working Memory. *BioRxiv*, 1–38. https://doi.org/10.1101/381848

Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall?

Cognitive Psychology. https://doi.org/10.1016/j.cogpsych.2018.09.002

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15*(3), 535–542.

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, *49*(4), 446–468. https://doi.org/10.1016/S0749-596X(03)00095-0

Tables

Table 1 Sample Description (means (and standard deviations)) of Experiment 1 and 2

Experiment	Age Group	Age	vocabulary	digit-symbol task	calibrated pres. Time in ms
1	Younger	22.46(3.05)	29.81 (2.24)	67.48 (8.27)	481.25 (145.66)
	Older	70.9 (4.3)	32.38 (1.63)	46.5 (8.19)	604.49 (113.35)
	$PD_{age\text{-}effect}$	-	100% < 0 < 0%	0 % < 0 < 100%	99.8% < 0 < 0.2%
2	Younger	24.67 (3.15)	27.64 (4.86)	66.71 (8.89)	462.5 (134.23)
	Older	70.76 (3.65)	33.81 (7.57)	49.05 (9.49)	656.83 (115.29)
	$PD_{age\text{-}effect}$	-	99.3% < 0 < 0.7%	0 % < 0 < 100%	100% < 0 < 0%

Note. The posterior density (PD) of the age effects. Zero represents the point of no age differences, and the percentages indicate how much of the estimated effect's posterior distribution lies below and above 0. Values below 0 reflect an advantage/higher values of older adults whereas positive values indicate a younger adults' advantage.

Table 2 The posterior effect estimates of the pairwise contrasts and their 95 % HDRs of the generalized linear mixed model for the immediate serial memory data of Experiments 1.

contrast	age group	Mode of parameter on logit scale	95% HDR
long baseline vs. short baseline	**young**	**0.58**	**[0.18, 0.95]**
	old	**0.49**	**[0.15, 0.84]**
sentence imagery vs. long baseline	young	-0.27	[-0.73, 0.2]
	old	**-0.56**	**[-0.96, -0.15]**
sentence grammar vs. long baseline	**young**	**-0.64**	**[-1.05, -0.25]**
	old	**-0.52**	**[-0.88, -0.16]**
sentence grammar vs. sentence imagery	young	-0.39	[-0.78, 0.04]
	old	0.07	[-0.37, 0.41]
sentence grammar vs. short baseline	young	-0.07	[-0.46, 0.32]
	old	-0.02	[-0.39, 0.32]
sentence imagery vs. short baseline	young	0.3	[-0.2, 0.79]
	old	-0.06	[-0.54, 0.4]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 3 The posterior effect estimates of the pairwise contrasts and their 95 % HDRs of the generalized linear mixed model for the delayed memory data of Experiment 1.

contrast	age group	Mode of parameter on identity scale	95% HDR
long baseline vs. short baseline	**young**	**0.09**	**[0.04, 0.14]**
	old	0.02	[-0.04, 0.07]
long baseline vs. sentence imagery	**young**	**-0.17**	**[-0.24, -0.1]**
	old	-0.03	[-0.11, 0.04]
long baseline vs. sentence grammar	young	-0.04	[-0.11, 0.04]
	old	-0.02	[-0.09, 0.06]
sentence grammar vs. sentence imagery	**young**	**-0.14**	**[-0.2, -0.07]**
	old	-0.02	[-0.09, 0.05]
sentence grammar vs. short baseline	**young**	**0.13**	**[0.06, 0.18]**
	old	0.03	[-0.03, 0.1]
sentence imagery vs. short baseline	**young**	**0.26**	**[0.18, 0.34]**
	old	0.05	[-0.03, 0.13]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 4 The posterior effect estimates of the pairwise contrasts and their 95 % HDRs of the generalized linear mixed model for the immediate serial memory data of Experiments 2.

contrast	**age group**	**Mode of parameter on logit scale**	**95% HDR**
long baseline vs. short baseline	**young**	**0.64**	**[0.24, 1.03]**
	old	0.19	[-0.35 , 0.7]
sentence imagery vs. long baseline	young	-0.16	[-0.55 , 0.24]
	old	**-0.53**	**[-1.02 , -0.05]**
sentence grammar vs. long baseline	**young**	**-0.55**	**[-0.92 , -0.17]**
	old	**-0.82**	**[-1.29 , -0.36]**
sentence grammar vs. sentence imagery	**young**	**-0.38**	**[-0.77 , -0.01]**
	old	-0.29	[-0.74 , 0.15]
sentence grammar vs. short baseline	young	0.1	[-0.31 , 0.49]
	old	**-0.63**	**[-1.14 , -0.16]**
sentence imagery vs. short baseline	young	0.48	[-0.05 , 1.04]
	old	-0.35	[-0.97 , 0.25]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 5 The posterior effect estimates of the pairwise contrasts and their 95 % HDRs from the generalized linear mixed model for the delayed memory data of Experiment 2.

contrast	age group	Mode of parameter on identity scale	95% HDR
long baseline vs. short baseline	**young**	**0.08**	**[0.02, 0.13]**
	old	0.04	[-0.01, 0.1]
long baseline vs. sentence imagery	**young**	**-0.13**	**[-0.19, -0.06]**
	old	-0.02	[-0.08, 0.05]
long baseline vs. sentence grammar	young	-0.03	[-0.09, 0.03]
	old	-0.01	[-0.07, 0.05]
sentence grammar vs. sentence imagery	**young**	**-0.1**	**[-0.18, -0.02]**
	old	-0.01	[-0.09, 0.06]
sentence grammar vs. short baseline	**young**	**0.11**	**[0.04, 0.16]**
	old	0.04	[-0.01, 0.1]
sentence imagery vs. short baseline	**young**	**0.2**	**[0.14, 0.27]**
	old	0.06	[-0.01, 0.12]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 6 The posterior effect estimates of the interactions of the pairwise contrasts of conditions with memory load, and their 95 % HDRs from the generalized linear mixed model for the immediate serial memory data of Experiments 1 & 2.

contrast	**Mode of the effect of memory load**	**95% HDR**
long baseline vs. short baseline	0.20	[-0.22, 0.58]
sentence imagery vs. long baseline	-0.06	[-0.45, 0.37]
sentence grammar vs. long baseline	-0.20	[-0.58, 0.16]
sentence grammar vs. sentence imagery	-0.14	[-0.22, 0.53]
sentence grammar vs. short baseline	0.35	[0.00, 0.80]
sentence imagery vs. short baseline	0.23	[-0.29, 0.75]

Note. Credible effects, defined as HDRs excluding zero, are printed in bold.

Table 7 The posterior effect estimates of the pairwise contrasts and their 95 % HDRs of the generalized linear mixed model for the immediate serial memory data collapsed over Experiments 1 & 2.

contrast	age group	Mode of parameter on identity scale	95% HDR
long baseline vs. short baseline	**young**	**0.62**	**[0.33, 0.86]**
	old	**0.35**	**[0.05, 0.65]**
sentence imagery vs. long baseline	young	-0.21	[-0.49, 0.06]
	old	**-0.52**	**[-0.82, -0.23]**
sentence grammar vs. long baseline	**young**	**-0.57**	**[-0.84, -0.35]**
	old	**-0.68**	**[-0.95, -0.40]**
sentence grammar vs. sentence imagery	**young**	**-0.38**	**[-0.65, -0.10]**
	old	-0.16	[-0.44, 0.14]
sentence grammar vs. short baseline	young	0.03	[-0.28, 0.28]
	old	**-0.34**	**[-0.60, -0.03]**
sentence imagery vs. short baseline	**young**	**0.38**	**[0.03, 0.74]**
	old	-0.17	[-0.53, 0.19]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 8 The posterior effect estimates of the interactions of pairwise contrasts of conditions with the effect of the memory load, and their 95 % HDRs from the generalized linear mixed model for the delayed memory data of Experiments 1 & 2.

contrast	**Mode of the effect of memory load**	**95% HDR**
long baseline vs. short baseline	0.02	[-0.07, 0.04]
sentence imagery vs. long baseline	0.02	[-0.05, 0.08]
sentence grammar vs. long baseline	0	[-0.07, 0.06]
sentence grammar vs. sentence imagery	0.02	[-0.05, 0.09]
sentence grammar vs. short baseline	-0.02	[-0.08, 0.04]
sentence imagery vs. short baseline	-0.04	[-0.10, 0.04]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Table 9 The posterior effect estimates of the pairwise contrasts and their 95 % HDRs of the generalized linear mixed model for the delayed memory data collapsed over Experiments 1 & 2.

contrast	age group	Mode of parameter on identity scale	95% HDR
long baseline vs. short baseline	**young**	**0.08**	**[0.04, 0.12]**
	old	0.03	[-0.01, 0.07]
sentence imagery vs. long baseline	**young**	**-0.15**	**[-0.20, -0.10]**
	old	-0.02	[-0.07, 0.02]
sentence grammar vs. long baseline	young	-0.04	[-0.08, 0.02]
	old	-0.01	[-0.05, 0.04]
sentence grammar vs. sentence imagery	**young**	**-0.11**	**[-0.17, -0.07]**
	old	-0.02	[-0.07, 0.03]
sentence grammar vs. short baseline	**young**	**0.11**	**[0.08, 0.15]**
	old	0.04	[0.00, 0.08]
sentence imagery vs. short baseline	**young**	**0.23**	**[0.18, 0.28]**
	old	0.06	[0.01, 0.10]

Note. Credible differences, defined as HDRs excluding zero, are printed in bold.

Figures

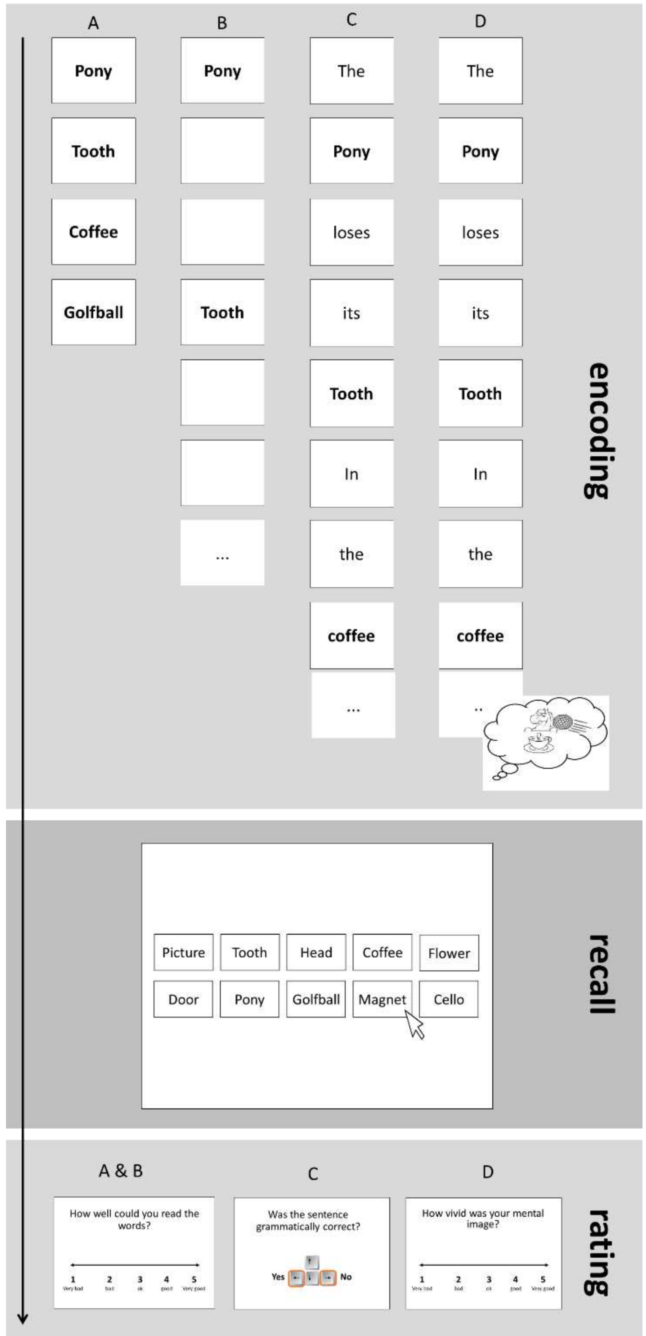

Figure 1 Illustration of the working memory paradigm of Experiment 1 and 2. Subjects were shown a list of words sequentially according to the four experimental conditions: A) the short baseline, B) the long baseline C) the sentence grammar condition and D) the sentence Imagery condition.

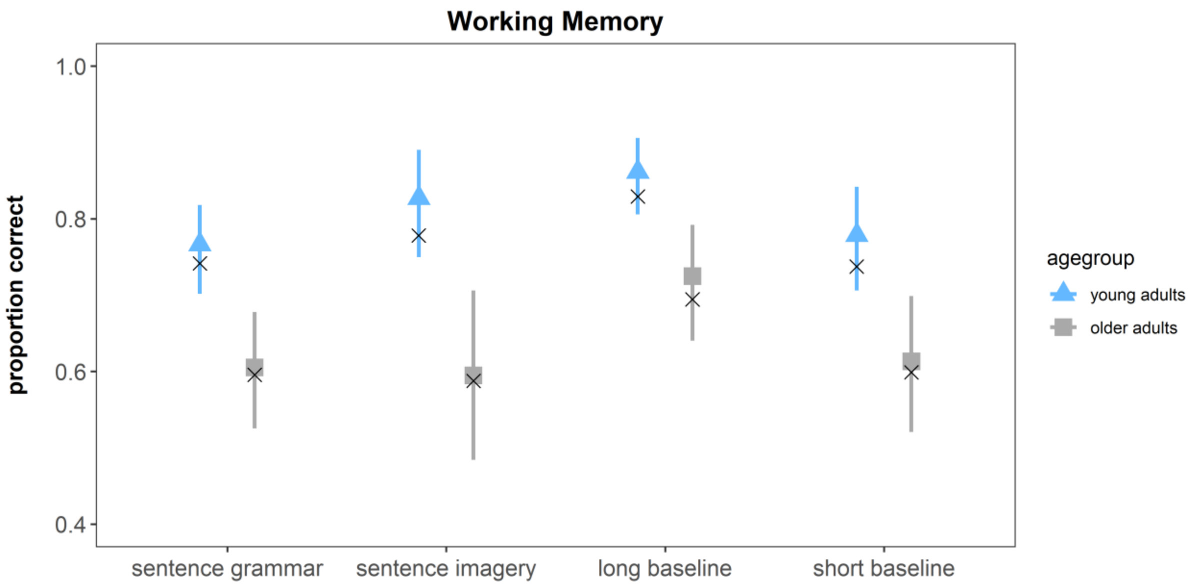

Figure 2 Proportion correct in the working-memory task in Experiment 1. The blue (young adults) and grey (older adults) symbols and error bars represent estimated proportions and their 95% HDRs from the BGLMM. The crosses represent the observed proportions. Their overlap indicates that the model adequately describes the data.

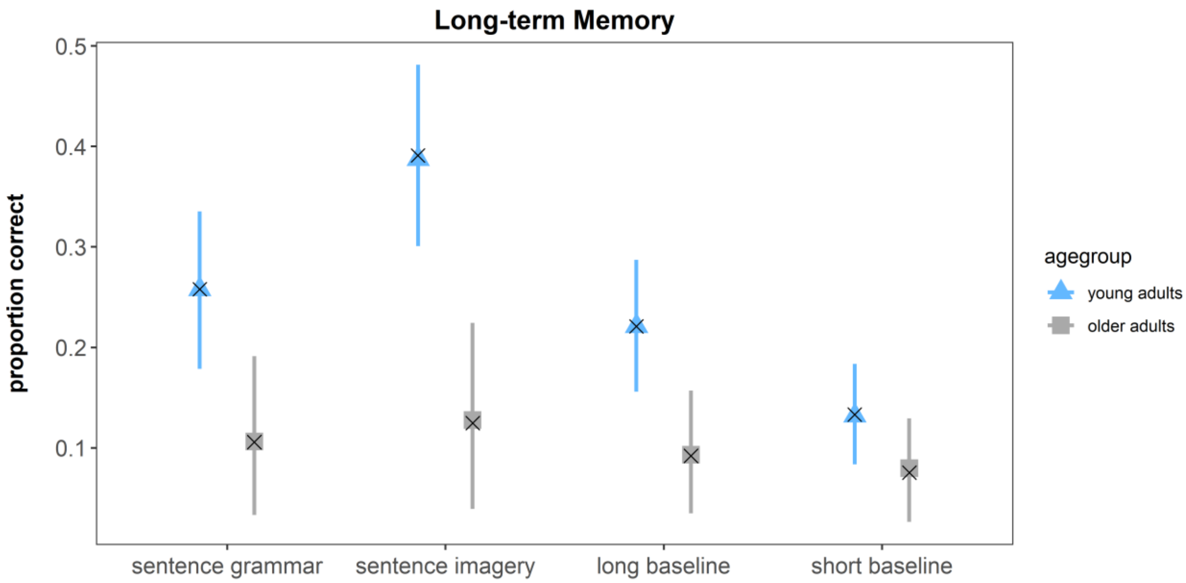

Figure 3 Proportion correct in the long-term memory task of Experiment 1 in t. The blue (young adults) and grey (older adults) symbols and error bars represent estimated proportions and their 95% HDRs from the BGLMM. The crosses represent the observed proportions. Their overlap indicates that the model adequately describes the data.

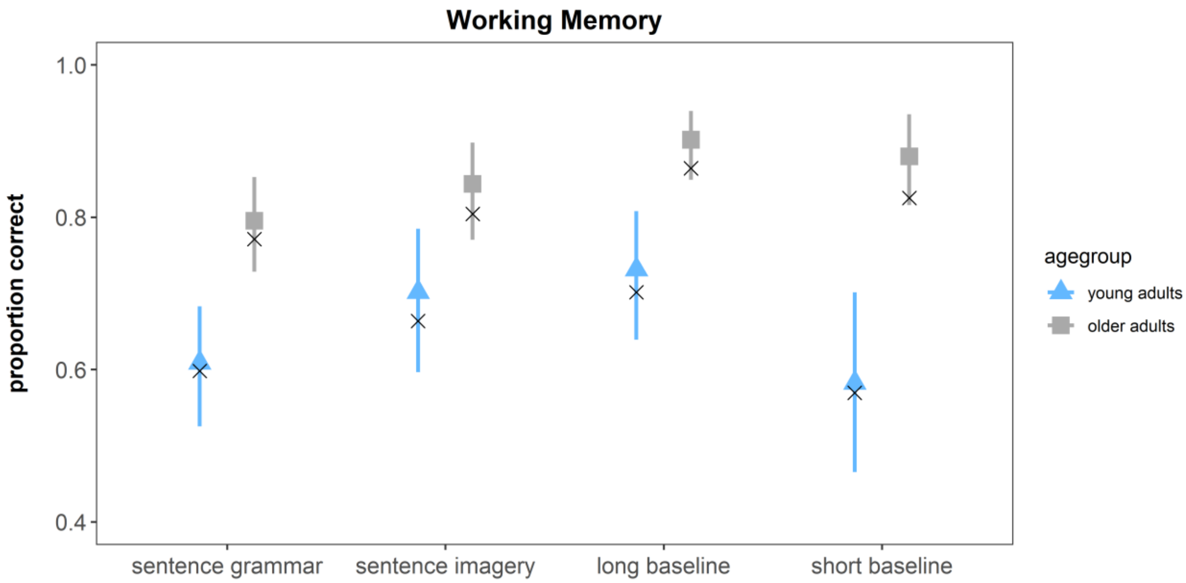

Figure 4 Proportion correct of the data from Experiment 2 in the working memory task. The blue (young adults) and grey (older adults) symbols and error bars represent estimated proportions and their 95% HDRs from the BGLMM. The crosses represent the observed proportions. Their overlap indicates that the model adequately describes the data.

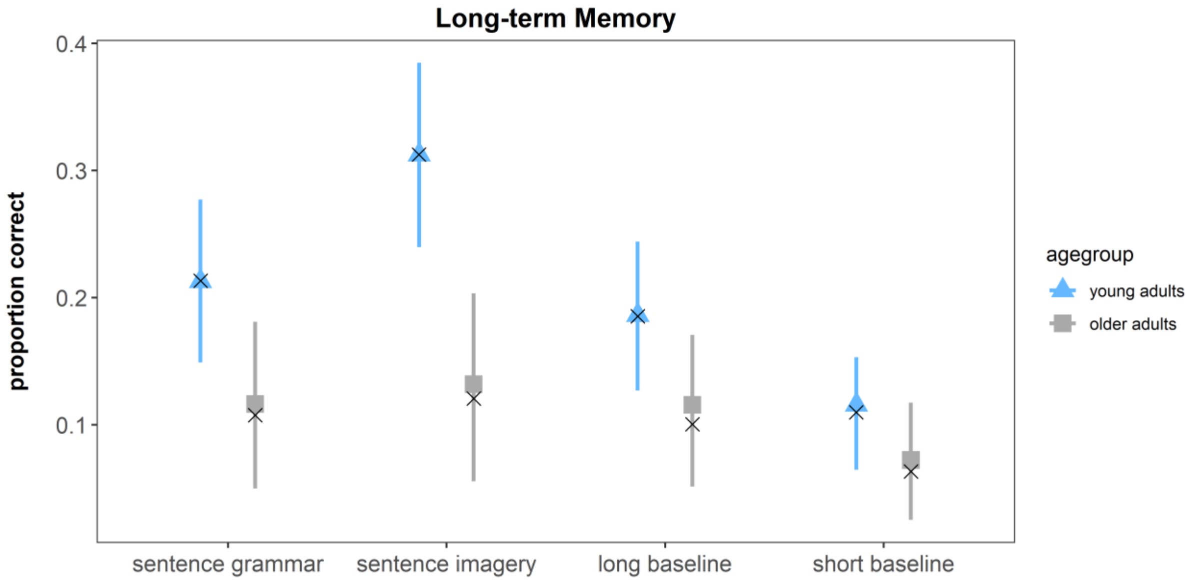

Figure 5 Proportion correct of the data from Experiment 2 in the long-term memory task. The blue (young adults) and grey (older adults) symbols and error bars represent estimated proportions and their 95% HDRs from the BGLMM. The crosses represent the observed proportions. Their overlap indicates that the model adequately describes the data.

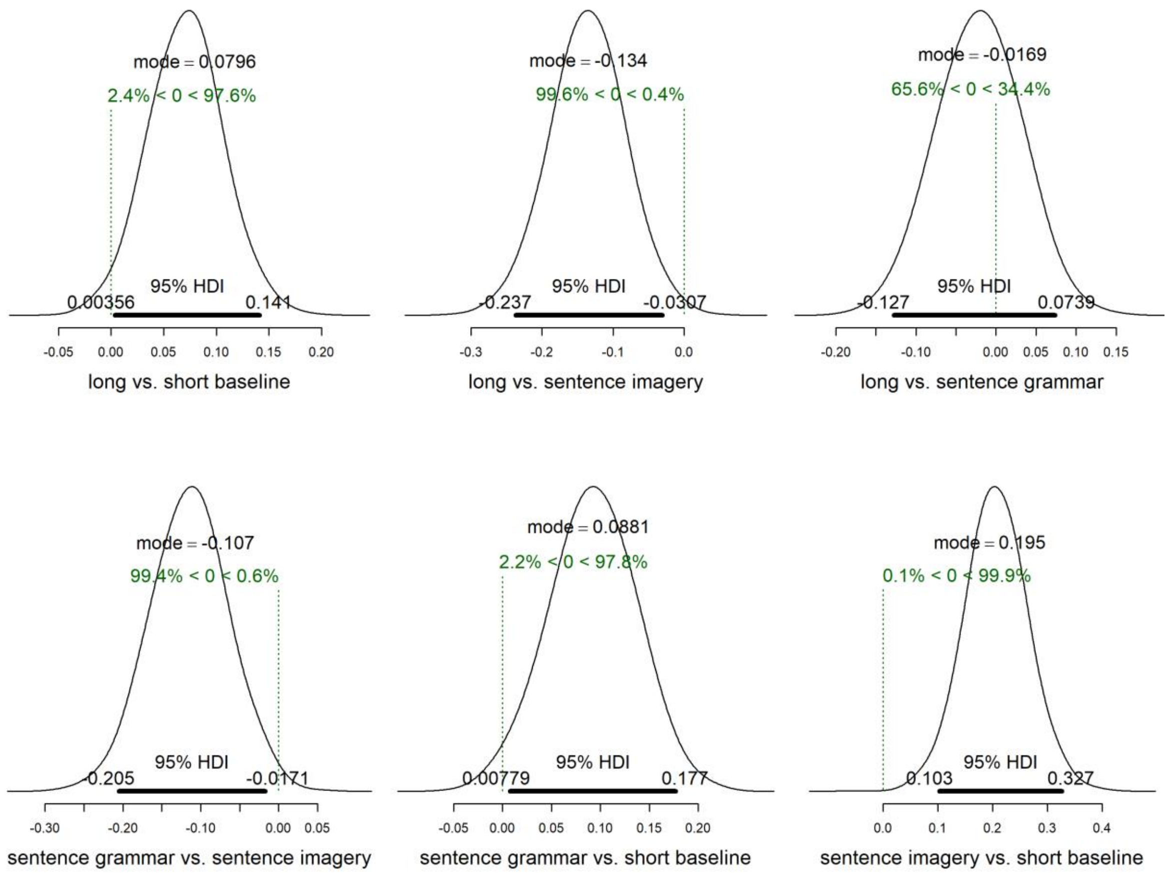

Figure 6 Posterior distributions of differences between the age groups in the effect of the respective conditions in LTM in Experiment 1. The mode and the highest density intervals reflect the effect size of the differences of pairwise condition contrasts between the age groups s. The dotted line indicates the point of no difference in the the condition contrast between the age groups. HDI's including zero reflect that there is no credible interaction of age with the condition contrast.

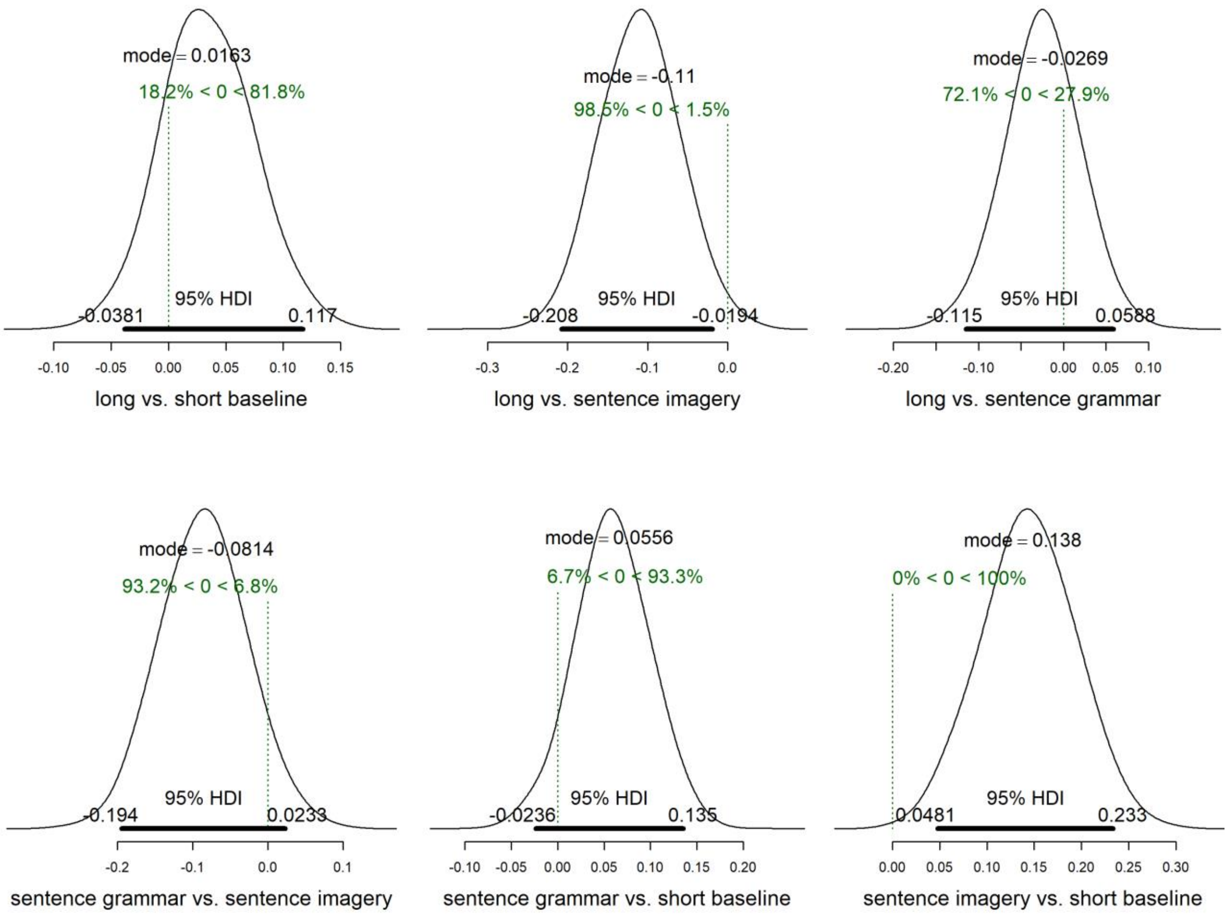

Figure 7 Posterior distributions of differences between the age groups in the effect of the respective conditions in LTM in Experiment 2. The mode and the highest density intervals reflect the effect size of differences of pairwise condition contrasts between the age groups. The dotted line indicates the point of no difference in the size of the condition contrast between the age groups. HDI's including zero reflect that there is no credible interaction of age with the condition contrast.